Exam 1 Review Questions (with answers)

1. A nutritionist is interested in whether hunger reduces the occurrence of smiling in construction workers. Twenty construction workers are selected from people working near the building where the nutritionist works. Ten were given a full lunch, while the other ten weren't given any food during the day. For 3 weeks, the researcher keeps records on the number of times the participants smiled during the workday. A mean for each group is then computed, using the number of smiles done daily, and these means are compared to determine whether hunger had an effect on smiling. In the experiment described here, state the independent variable, the dependent variable, the sample, the population, the data, and the statistic
    a. IV: given food or not given food
    b. DV: number of smiles during the day
    c. Sample: 20 construction workers from around the building
    d. Population: construction workers
    e. Data: amount of smiles each participant has over 3 weeks
    f. Statistic: the mean number of smiles of participants who were given a meal and the mean number of smiles of participants who weren't given a meal
2. You have a theory that your significant other's happiness affects your attraction to them. You operationalize attraction as your heart rate and your partner's happiness as how long they smile in your presence. You decide to measure your partner's happiness every time they are with you, and then measure your heart rate after each smile ends. The data you obtained is below. What is the predictor and criterion variable? What correlation coefficient should be used, and calculate it, and interpret it.
    a. Predictor: how long your partner smiles
    b. Criterion: your heart rate
    c. Correlation coefficient: Pearson's r because both variables are ratio level.
        i. Length of smile is ratio because 0 seconds means there was no smile
        ii. Heart rate is ratio because 0 bpm means literally no heartbeat
    d. r = 0.5475; There is a positive correlation between your partner's smile length and your heart rate. As smile length increases, so does your heart rate.

Smile	Length of Smile (in seconds)	Heart Rate	x^2	y^2	xy
1	35	86	1,225	7,396	3,010
2	20	99	400	9,801	1,980
3	43	104	1,849	10,816	4,472
4	10	85	100	7,225	850
5	43	98	1,849	9,604	4,214
Sum:	151	472	5,423	44,842	14,526

3. You observe which type of Girl Scout cookies people buy at the table and get the following data: Thin Mints, Trefoils, Trefoils, Samoas, Samoas, Thin Mints, Samoas, Thin Mints, Do-Si-Dos. Which measure of central tendency would you use to describe the data?
    a. Mode
4. You have a data set with σ= 3.2 and μ= 83. You transform all of your data to z-scores. With your z-score distribution, what is your mean and standard deviation?
    a. μ= 0, σ= 1
5. Fill the chart with the appropriate symbols:

	Population	Sample
Mean	μ	$\bar{x}$
Standard Deviation	σ	s
Variance	σ^2	s^2
Denominator of Variance Formula	N	n-1

6. You want to see what the temperature is in all of your friends' houses. You obtain all of their temperatures (in Fahrenheit): 93, 65, 82, 75, 82. What level of measurement is this? Calculate the three measures of central tendency, and state which one you would report. Calculate the standard deviation and variance and interpret what they mean.
    a. Level of measurement: Interval
    b. Mean: 79.4, Median: 82, Mode: 82
        i. Report the mean because there are no extremes, and it uses all of the data
    c. Variance: 85.04, SD: 9.22 (population because it is ALL of your friends)
        i. The variance means that, on average, our data are 85.04 squared units away from the mean.
        ii. The standard deviation means that, on average, our data is 9.22 units away from the mean.
7. You want to take a sample of 13 students from your Communications class of 124 students, so you decide to use the first 13 students that walk into the class. What sampling method is this?
    a. Convenience sampling
8. A predictor variable is always paired with a ___Criterion variable____

9. You have a set of exam scores with σ= 5, μ= 63, N= 40. You realize that there are questions all of the students got wrong and decide to add 6 points to each score. What is your new standard deviation?
    a. σ = 5
10. You gathered exam scores from 7 friends and found their scores to be 96, 61, 76, 84, 62, 86, 92. Create a grouped frequency table with this data, using 4 intervals. Your lowest level of measurement is 1.
    a. Range: 96-61=35
        i. Interval width = 35/4 = 8.75 ~ 9

Apparent Limits	Real Limits	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency	Percentile
61-69	60.5-69.5	65	2	.29	2	.29	29%
70-78	69.5-78.5	74	1	.14	3	.43	43%
79-87	78.5-87.5	83	2	.29	5	.72	72%
88-96	87.5-96.5	92	2	.29	7	1.00*	100%

Because of rounding error, we get 1.01, but we round down to 1.00

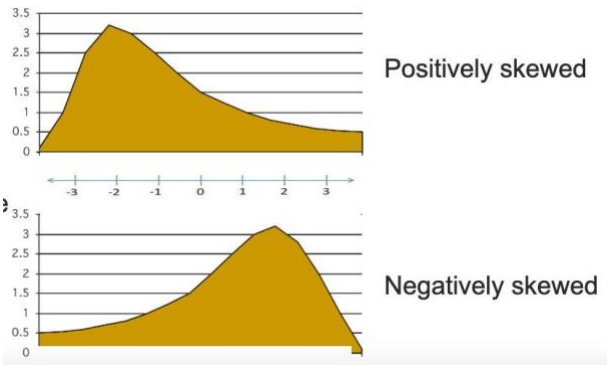
11. From the previous problem, would you use a histogram or bar graph, and why?
    a. A histogram because histograms are used for interval or ratio data, and exam scores are ratio.
12. You want to sample universities and colleges in Maryland, so you decide to separate them into cities, and randomly sample 7 of the cities as a whole. Which sampling method are you using?
    a. Cluster sampling
13. What is the purpose of z-scores?
    a. To standardize data so that comparisons can be made
14. Data for inferential statistics is collected from a __sample_____ and theoretically generalizable to a ___population____.
15. Imagine that you are conducting a study and one of the variables you want to examine is levels of depression. Depending how you define levels of depression, this variable could be categorized as continuous OR discrete.
    a. Give an example of how depression scores could be measured in a way that is discrete
        i. Imagine you have a depression scale that ranges from 0-20. If you decide to define levels of depression as “low” (scoring between 0-10), “moderate” (10-15), and “severe” (15-20). Then, they are discrete. You

can't have somebody who is between low and moderate, or between moderate and severe. Every person in your study is only one of the three levels.

   b. Give an example of how depression scores could be measured in a way that is continuous
      i. Imagine you have the same depression scale, and you ask participants to rate themselves on a continuous sliding bar. That is, participants can give an answer anywhere on the scale between 0-20. They could score 4, 7.98222, 10.8, 15.82222, and anywhere in between. This is continuous because participants can score anywhere on the scale.

16. What's the importance of random assignment?
    a. Random assignment hopefully distributes any potential individual differences between people/groups in your study in such a way that these differences no longer have a significant effect on the results.
17. Imagine you and your friend want to study learning strategies. You decide to observe students' studying in the library. Your friend decides to directly ask students about their study habits as well as their recent exam scores, and then see how related they are.
    a. Which of you did a correlational study and which did a naturalistic study? Can you determine cause and effect from either of these study types?
       i. You did the naturalistic study, and your friend did the correlational study. You can't determine cause and effect from either of these studies because nothing was manipulated by the researchers. Experimental studies are the only type that can determine cause and effect.
18. List the four levels of variables from least information to most information
    a. (least) Nominal - ordinal - interval - ratio (most)
19. With what type of data is it most appropriate to use the mode to describe central tendency?
    a. Nominal data
20. In a standard, normal distribution curve, what is the relationship between the median, mean, and mode?
    a. Mean = median = mode
    b. How does the answer change if you have a positively skewed distribution? A negatively skewed one?
       i. Positively skewed: mode < median < mean
       ii. Negatively skewed: mean < median < mode
21. Explain (in your own words) why the mean, median and mode shift depending on the skew of the distribution
    a. The mean is most sensitive, will always get pulled towards the tail. For the mode, the most frequent score will not change due to the addition of outliers. The median gets pulled a little, but not as much as the mean does because the median isn't as sensitive as the mean is.

22. Draw the general shape of a positively skewed distribution and a negatively skewed one.
    a. See pictures to the left



23. What is **always** the sum of all the deviations of scores around the mean?
    a. 0
24. What's the point of taking the square root of the variance?
    a. Gets you back to your original unit of measurement
25. What's the name of a distribution that is entirely z scores rather than raw scores?
    a. The standard distribution
        i. When the distribution of the original scores is normally distributed, It is called the standard normal distribution.
26. Pick **one** of the following topics

- correlation
- z-scores
- standard deviation
- central tendency

Write how you would explain the concept to each of the following audiences:

- An elementary school student
- A middle school student
- A high school student who is enrolled in an intro level statistics course
- A fellow classmate in PSYC200
- A TA/professor from PSYC200 (here you should include an example where you're taking the concept and applying it to something)

Answers may vary, but the idea is that you build on the concept at each level. To answer this questions, you should be able to explain what the concept is at an extremely simplified level (i.e., we have measurements that tell us how the majority of people respond to something) all the way to the level we discussed in class (i.e., there are multiple ways to measure central tendency such as mean, median and mode. Different types of data determine which measure of central tendency is most appropriate. For example, if someone was studying the relationship between height and the MD county they live in, they might use mode to describe central tendency since the county is a nominal variable. If you were studying how long (in miles) someone's commute was and the amount of sleep they receive, you might use mean to describe the central tendency of commute length, since miles is a ratio level measurement.)